



Målingsbasert Prosessforbedring

*The road to wisdom?
Well, it's plain and simple to
express:
Err and err and err again
but less and less and less.
(Kumbel/ Piet Hein)*



Målingsbasert prosessforbedring er vanskelig ...

- Problemet er ofte å klare å måle det vi virkelig er ute etter å måle:
 - kvalitet
 - effektivitet
 - brukervennlighet
 - Vedlikeholdbarhet
 - ...
- Det er så mange variable vi ikke klarer å holde konstante (ref. statistisk prosesskontroll).
- Systemutvikling er en menneskebasert aktivitet der de individuelle ferdighetene varierer svært mye.
- I foregående studium ble antall feil brukt som målestørrelse for kvalitet. Er det meningsfullt?
- **Øvelse:** Hva menes med: «Det du ikke kan måle, forstår du ikke» - «Det du ikke forstår, kan du ikke måle»?
- **"Konklusjon":** Vi trenger litt måleteori.

INF5180

3



Hvorfor målinger?

- For å få mer **objektiv** kunnskap
 - *Fra*
☞ "jeg tror at antall feil i programvaren vi produserer er blitt lavere de siste årene"
 - *til*
☞ "feil per 1000 linjer kode funnet i akseptansetest har blitt redusert fra 3 til 1"
- For å kunne analysere årsakssammenhenger og dermed **lære** av erfaringer
 - *Analyser kan f.eks vise om nye rutiner (som f.eks. parprogrammering) har hatt positiv effekt og muliggjøre en prediksjon av kvalitetsaspekter*
- For å kunne **validere** om målsetninger er oppfylt
 - *Målbarhet i kvalitetsrelaterte krav tvinger oppdragsgiver til å angi kravene så presist som mulig. Krav som ikke er "falsifiserbare" har ofte liten verdi.*

INF5180

4



Måleteori - definisjon av målestørrelse

- **Entitet:** En ting i den "virkelige" verden. (Egentlig, vår modell av en ting i den virkelige verden!) som for eksempel en programvaremodul eller et prosjekt.
- **Attributt:** En egenskap til en entitet. Eksempel: Lengde/størrelse av en programvaremodul
- **Empirisk relasjonelt system:** $\langle E, (R1..Rn) \rangle$, hvor E er en mengde entiteter og $R1..Rn$ er relasjoner mellom entitetene i E mhp på gitte attributter.
 - Eksempel: $E = \{\text{modul1}, \text{modul2}\}$, $R1 = \langle \text{modul1.lengde} > \text{modul2.lengde} \rangle$
- **Numerisk relasjonelt system:** $\langle N, S1..Sn \rangle$, hvor N er en mengde tall eller symboler og $S1..Sn$ er relasjoner mellom tallene/symbolene i E.
 - Eksempel: $N = \{\text{alle reelle tall}\}$ $S1 = \langle 445 > 400 \rangle$



Hvorfor måleteori?

- For å forstå hva som skjer når vi måler.
- For å vite hvilke målinger (i systemutvikling) som er meningsfulle.
 - Lengdemålinger synes meningsfullt, kvalitetsmålinger noe mer komplisert, og målinger av intelligens tvilsomt. Hvorfor?
- For å vite hvilke operasjoner og kalkulasjoner som er meningsfulle på de målte verdiene.
 - Når er utregning av aritmetisk middel meningsfullt?
- Et instrument til å vurdere hva som er meningsfulle kvalitetsutsagn basert på målingene, dvs for å kunne interpretere måleresultatene!
- Hvorfor er det meningsfullt å si at temperaturen steg med 5 grader Celsius, men ikke med 10%?
- For å kunne vurdere målestørrelser.



Ulike skalaer

- **Nominalskala**
 - brukes til klassifisering i gjensidig ekskluderende klasser
 - Eks: Rød, hvit, blå etc
- **Ordinalskala**
 - Ordnet nominalskala
 - Eks: Medium, stor, ekstra stor (pizza)
 - Eks: Nivå 1 til 5 i CMM-modellen
- **Intervallskala**
 - Brukes når likhet, ordning og forskjell mellom verdier gir mening, men ikke prosentvis forskjell (som forlanger et nullnivå).
 - Eks: Temperatur: Grader Celsius kan transformeres til Farenheit via en formel
- **Rationalskala**
 - Intervallskala med fast, veldefinert nullpunkt
 - Eks: Antall kodelinjer
 - Eks: Temperatur i Kelvin

INF5180

7



Måleprogram - Motivering

«In practice, there are very good reasons why software development staff view metrication programmes as, at best, an unwelcome overhead and, at worst, a major threat.»

Øvelse: Hvilke grunner?

- => Hvordan motivere til engasjement og bidrag?
- konsultasjon og involvering
 - brukervennlig datainnsamlings-systemer
 - opplæring
 - tilbakemelding

INF5180

8



Måleprogram- Forankring

Topplederen:

- målinger forankret i ledelsens målsetninger (strategiske mål, kontroll over prosjekter, bygg på ledelsens "kjepphester"...)
- stimuler til at strategiske målsetninger formuleres (og følges opp) kvantitativt
- la målinger inngå som en integrert del av kvalitetssystem og prosesser

Mellomlederne:

- "Eier" de konkrete målingene og målsetningene
- Må stimuleres til å konkretisere de overordnede målene
- Kan lett oppleve forbedringsprogram og målsetninger som en trussel

Grunnplanet:

- Enkelte personer har ofte overraskende stor innflytelse
- De mest innflytelsesrike er ofte tilfreds med situasjonen og kan oppleve endringer som en trussel
- Ekte engasjement fra enkeltpersonene er en forutsetning for å lykkes

INF5180

9



Måleprogram- Planlegging

Elementer i planen:

- Målsetninger
 - *f eks vha GQM*
- Datainnsamlingssystem
 - *definering av alle målestørrelser*
 - *definering av måleprosess*
 - *spesifikasjon av rapporter*
- Installering/opplæring
 - *opplæring i datainnsamling*
 - *installering av verktøy m.m.*
- Operasjon
 - *datainnsamling, analyse og rapportering*
- Vedlikehold/Avslutning
 - *forbedring, modifikasjoner av måleprogram*
 - *avslutning dersom ikke behov lenger er tilstede*

INF5180

10



Måleprogram- Organisering

Anbefales ofte å opprette et «Måle-team» med ansvar for:

- innføring og vedlikehold av måleprogrammet som helhet
- nødvendige endringer i prosedyrer og standarder for å gjennomføre måleprogrammet
- datainnsamling, validering og lagring av prosjekt-data
- trend-analyse, inkludert utvikling av kostnadsestimeringsmodeller og langsiktig overvåking av produktivitet og kvalitet
- bruk av målinger i prosessforbedring

Øvelse: Denne anbefalingen innebærer en sentralisering av måle/analyse-ansvaret. Hvorfor ikke distribuere det på samme måte som med kvalitet i TQM-rammeverket ("kvalitet er alles ansvar")?



Måleprogram- Vedlikehold

Et måleprogram bør ha prosedyrer for å:

- lagre praktiske problemer med måleprogrammet (egen læring og forbedring)
- ha regelmessige gjennomganger av måleprogrammet
- innføre endringer i datainnsamlingsrutinene ved behov



Datainnsamling - kriterier for nyttige data

- Anbefalt å spørre følgende fire spørsmål:
 - forteller det deg hva som skjer?
 - forteller det deg hva som er riktig eller galt?
 - forteller det deg det du ønsker å vite?
 - forteller det deg det du trenger å vite?
- Faktorer som avgjør nytte:
 - relevans
 - «timeliness» (kommer målingen f eks i tide for å påvirke beslutninger)
 - repeterbarhet/sammenlignbarhet
 - nøyaktighet/fullstendighet

INF5180

13



Datainnsamling

- Verifikasjon
 - Er alle data innhentet?
 - Er data innbyrdes konsistente?
 - Synes dataene å være plausible?
- Analyse og rapportering
 - måleprogrammets godhet bestemmes av hvilke analyser og rapporter som kan produseres!!!
 - type analyser og rapportering bestemmer hva slags datainnsamling som behøves (tilfeldig utvalg, eksperiment,) og hvor ofte data må innhentes
- Integrasjon
 - datainnsamling og tilbakemelding bør være en integrert del av utviklings/vedlikeholdsprosessene
 - datainnsamling bør, så langt som mulig, benytte seg av eksisterende registreringsrutiner
- Automatisering
 - Dersom det er mulig bør datainnsamling automatiseres (datainnsamling er stort sett lavstatus, kjedelig arbeid)
 - Erfaringsmessig er automatisk innsamlet data av begrenset verdi
 - Erfaringsdatabaser som gir automatiserte rapporter og gjør tidligere datainnsamling tilgjengelig for nye prosjekter kan ha stor verdi

INF5180

14



Måle på "kvalitet"?

- Vanlig utsagn: "Kvalitet er viktigst for oss".
 - Hvis det er så viktig:
 - ☒ Hvordan definerer dere så kvalitet?
 - ☒ Hvordan dokumenterer dere kvaliteten?
 - ☒ Hvordan følger dere opp kvaliteten i det som leveres/produseres?
- Kvalitet brukes som regel i forståelsen "Funksjonell kvalitet" (se del I) dvs at det virker etter hensikten.
- Ordet kvalitet brukes i dagligtalen i betydningen "holdbar" eller "pålitelig".
- Pålitelighet defineres ofte som **antall feil/tidsenhet** (MTBF) eller **antall feil / produktstørrelse**.
- Hvis vi har prosessforbedring som fokus er det nærliggende å tenke på "prosjektnivå" dvs **antall feil / produsert kodelinje i prosjektet**.

INF5180

15



Ytelsesmålinger (Tom Gilb*)

Performance	Effect of Change In Performance	Scale of Measure
<u>Customer Satisfaction</u>	Fewer letters of complaint	Number of letters complaining about a defined [Product] received within a defined [Time Period]
<u>Customer Satisfaction</u>	Fewer returned goods	Percentage of defined [Product] returned within defined [Time Period after Purchase] with defined [Customer Issue]
<u>Environment-ally Friendly</u>	Improved rating as measured on international standard	Number of defined [Product Type] failing defined [Test] within a defined [Time Period]
<u>User-friendly</u>	Fewer errors made	Percentage of defined [Transaction Type] with defined [Error] input by defined [User Type]
<u>User-friendly</u>	Faster time for completion of transactions	Time in minutes for a defined [Transaction] to be carried out to <satisfactory> completion

*se www.gilb.com
for en hel del stoff
om måling



Kvalitetsattributter ISO/IEC 9126

- **Functionality**
 - Overenstemmelse med krav / forventninger etc
- **Usability**
 - Brukeropplevelse, intuitivt UI etc
- **Maintainability**
 - Endringsvennlighet, utvidbarhet, testbarhet
- **Reliability**
 - Stabilitet, modenhet, feiltoleranse, robusthet
- **Efficiency**
 - Tids- og ressurs-forbruk
- **Portability**
 - Adapterbarhet, installerbarhet etc...

INF5180

17



Feil/LOC - enkel måleteoretisk analyse

Empirisk relasjonelt system:

- Entitet = programvare (eller programvaremodul)
- Attributter av interesse = "feil", "lengde på programmet" og "kvalitet"
- Relasjoner/operasjoner av interesse er bl a: "antall enheter feil", "lengde", "flere feil enn", "lengre enn" og "bedre kvalitet enn".
- Forståelse av feil: "enhver tilstand som atskiller seg fra det forventede" (IEEE 1044-1993)
- Forståelse av lengde: antall linjer kode (krever en språkavhengig tellestandard)
- Forståelse av kvalitet = antall feil normert for lengden
- NB: En forbedring av målestørrelsen kan innebære en klassifisering av feil mhp alvorlighet, innvirkning på kvalitet, innvirkning på kundefornøydhhet etc.

INF5180

18



Feil/LOC (forts.)

Numerisk relasjonelt system:

- **Entitet:** alle reelle tall (R)
- **Attributterrelasjoner:** ">"

Målestørrelser:

- Mapping fra antall feil og fra lengde til reelle tall. Dette er åpenbart en målestørrelse hvor ekvivalensen er oppfylt mhp relasjonene "flere feil enn" og "lengre enn". Skalatype for F (feil) og LOC (lengde) er begge ratio.

Prediksjon:

- Da det er vanskelig å etablere en fullstendig forbindelse mellom Feil/LOC og intuitiv forståelse av kvalitet, er Feil/LOC formelt sett kun en **prediksjon** av kvalitet. **NB:** Dette gjelder trolig også alle andre "målinger" av kvalitet.



Feil/LOC (forts.)

Problemer med $M(\text{programvare.feil} / \text{programvare.lengde})$ som prediksjonsmål på programvarekvalitet:

- "Beskriver" ikke en generelt akseptert forståelse av kvalitet.
- En økende F/LOC behøver ikke bety at kvaliteten minker. F eks vil en fjerning av redundans i programvaren redusere LOC uten at antall feil behøver å berøres. Mange vil hevde at kvaliteten øker pga lavere redundansgrad!. Feil/funksjonspoeng "løser" dette problemet, men innfører nye problemer.
- Hva eller hvem avgjør om noe er forventet eller ikke? Hva slags feil skal registreres? Funnet i akseptanse-test, meldt 1. måned etter leveranse, prediktert gjenstående, funnet av kunde? etc.?)



Feil/LOC (forts.)

- **Vurdering:** Fra et måleteoretisk synspunkt er feil/LOC en relativt "sunn" målestørrelse. Aktsomhet må utvises ved interpretasjonen feil/LOC = kvalitet. Denne målestørrelsen krever god innsikt i empirisk relasjonelt system og måleprosess for å kunne brukes fornuftig.



Hva med kostnader?

- Timeforbruk i prosjekt
 - Inkluder overtid, ekskluder linjeaktiviteter som avdelingsmøter etc
 - Skille på produktiv tid og uproduktiv tid?
 - Skille på feilretting, endringshåndtering og "ren utvikling"?
 - Fordeling av timeforbruk over fase / inkrement?
- Nødvendig opplæringskostnader
 - Fulle kompetansegap for å gjennomføre prosjektet
- Verktøykostnader
 - Ren innkjøp og evt løpende kostnader
 - Opplæringskostnader
 - Læringskurvekostnader?

MERK: For å kunne vise utviklingen (forbedring) på kostnadssiden må det relateres til produsert mengde produkt m.a.o. **produktivitet**.



Hva med tid?

- "Time-to-market er det desidert viktigste for oss" er et vanlig utsagn
 - Hvis det er så viktig:
 - ☒ Hvordan definerer du "time-to-market"?
 - ☒ Hvordan følger du med på utviklingen av denne parameteren?
 - Kalendertid må defineres
 - Antall arbeidsdager, antall uker, måneder osv
 - Krever at prosjektene/inkrementene har et klart definert starttidspunkt
 - Krever at prosjektene/inkrementene har et klart definert sluttidspunkt
- MERK: Det kan være et viktig, eksplisitt mål å makte flere releaser per år. Men for å forsikre seg om at omleggingen ikke fører til lavere produktivitet eller kvalitet må man også her inkludere **størrelse på produktet**.



Hva skal man velge å måle på?

- Det er uoverkommelig å måle på alt, det er i hvert fall svært vanskelig å forsvare ut fra en kost/nytte vurdering.
- Mye tyder på at **størrelse** er fellesnevneren som gir nytte i de fleste ulike målingene.
- Størrelse er overraskende vanskelig å måle.
 - Uansett må man ha streng definisjon av antall produserte kodelinjer, antall modifiserte kodelinjer og antall fjernede kodelinjer.
 - Må vi også telle antall *gjennbrakte* kodelinjer?
- Størrelse kan også måles på mer abstrakt, funksjonelt nivå. Her er **funksjonspunkter** (Function points) den mest kjente. Dette er en svært avansert teknikk som krever høyt generelt modenhetsnivå og stabilitet i krav og prosess. Av andre mulige størrelsesmål kan vi nevne antall Use-Cases, antall skjemaer, antall skjemaer eller antall tabeller i databasen.



Øvelsesoppgave

I systemutviklingsorganisasjonen "Din IT-parner AS" har til nå alle systemutviklingsprosessene vært beskrevet i en papirbasert håndbok. Nylig er det blitt utviklet en web-basert utgave av beskrivelsene, der det i tillegg er lagt inn "lenker" mellom relaterte dokumenter. Mao, mens papir-håndboken var sekvensiell har web-utgaven en nettverksstruktur .

IT-sjefen var i grunnen svært fornøyd med de papirbaserte håndbøkene og vil gjerne at det gjøres en empirisk sammenligning før de eventuelt går helt over til web-basert utgave.

Lag en plan for et måleprogram i organisasjonen. Måleprogrammet skal ha som mål å avgjøre hvilke av de to utgavene som er mest effektiv for organisasjonen.



Crosbys kvalitetskostnader

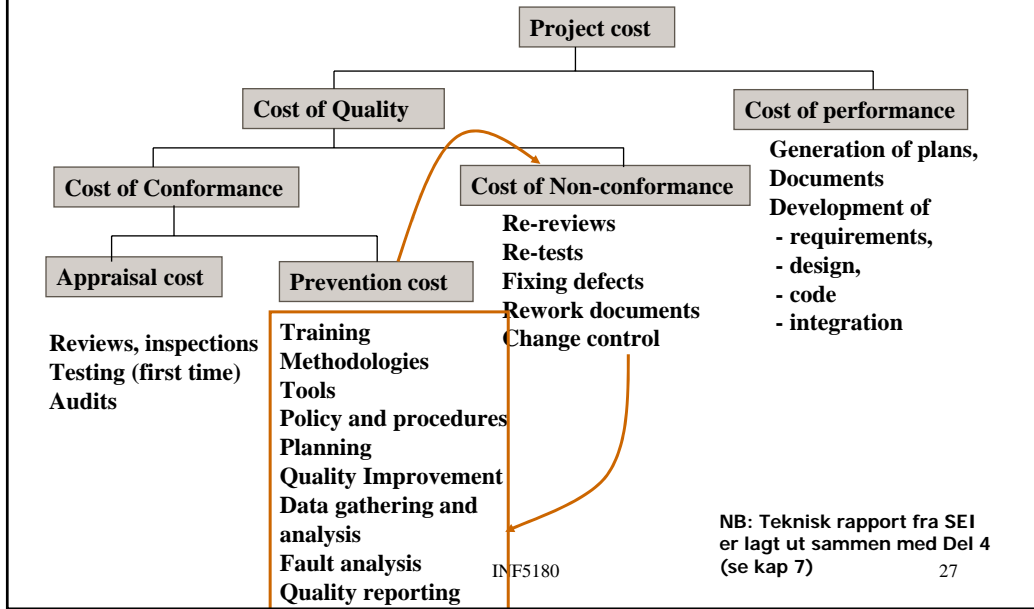
- Crosby definerer kvalitet som "conformance to requirements."
- Kvalitetskostnader har 3 komponenter:
 - Feilkostnader ("Failure cost"): det det koster å finne og rette en feil pluss det det koster å komme i drift igjen.
 - Evalueringskostnader ("Appraisal cost"): det det koster å evaluere produktet for å finne kvalitetsnivået
 - Forebyggende kostnader ("Prevention cost"): det det koster å identifisere årsakene til feil og å forhindre liknende feil å opptre i framtiden.

Referanse:

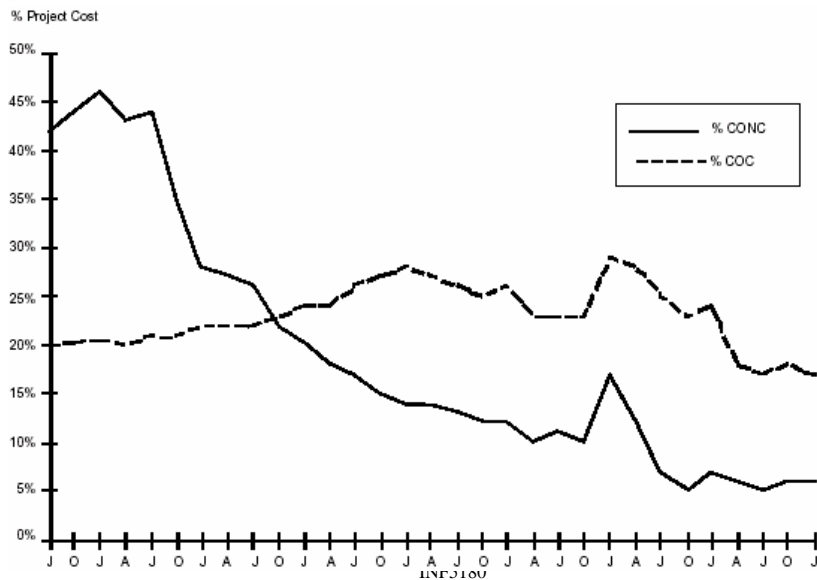
[Crosby] Philip B. Crosby, *Quality is Free, The Art of Making quality Certain*. New York: Mentor, New American Library, 1979.



Bruk av Crosbys modell i Raytheon



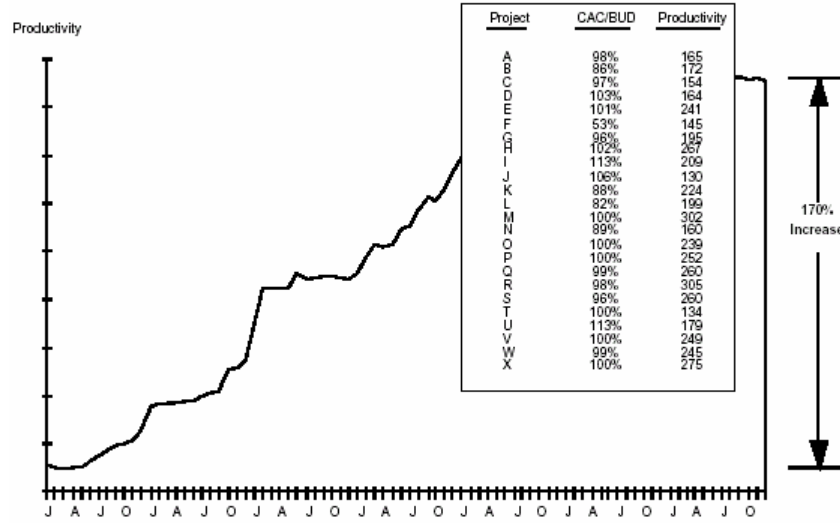
"Conformance"-utviklingen over 6 år



28



Produktivitetsøkning over 6 år

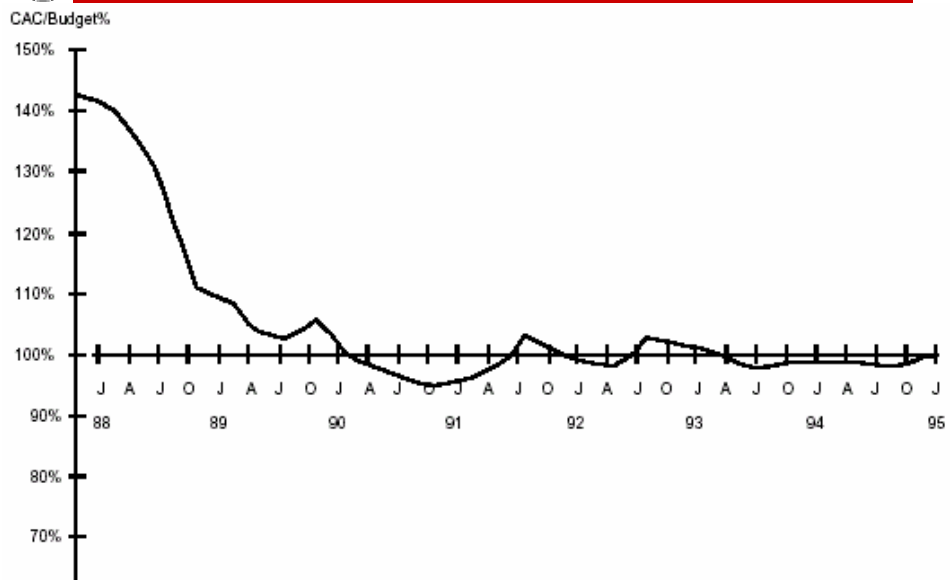


INF5180

29



Forutsigbarhet i prosjekter (7 år)



INF5180

30



Feiltetthet (over 7 år)



INF5180

31



Modellbasert prosessforbedring

- Den rene målingsbaserte filosofien baserer seg på at alle organisasjoner er forskjellige og har ulike utfordringer. Derfor finnes det ingen norm for hvordan ting skal gjøres – og det er unødvendig å forsøke å finne denne.
- På den andre siden så finnes et ganske stort antall av normative modeller eller rammeverk for prosessforbedring. Opphavsmennene til disse vil hevde at erfaring viser at det er en hel rekke felles problemer og utfordringer som må løses. Derfor er det fornuftig å beskrive retningslinjer for hvordan man kan gå fram for å forbedre seg.
- Dessuten har det vist seg i praksis at svært få lykkes over lengre tid med å kun basere seg på å måle og samle erfaring. Dette har i enkelte tilfeller ført til at store firmaer har laget sin egen modell for prosessforbedringsarbeidet...
- De fleste normative modellene er forholdsvis åpne i den forstand at de forteller *hva* som skal være på plass men ikke *hvordan* det skal løses.

INF5180

32



...



© Scott Adams, Inc./Dist. by UFS, Inc.

INF5180

33



Noen opprinnelige standarder

- British standard BS5750 i 1979 ledet etter hvert til ISO 9000 som kom ut i 1987. ISO 9000 er revidert i 1994 og 2000.
- Six Sigma er et kvalitetssystem som baseres på standardiserte målinger og som krever høy grad av repeterbarhet ved at det brukes statistiske teknikker på dataene.
- Balanced Scorecard fra 1996 fokuserer på å balansere ulike aspekter ved en virksomhet som bedriver utvikling av tjenester eller produkter.
- EFQM oppr. Fra 1988 var et initiativ fra 14 større europeiske selskaper som ønsket å fremme produktivitetsøkning i Europa ved å tildele en årlig pris til den beste organisasjonen i Europa.
- CMM (Capability Maturity Model) kom først ut i 1989 og er senere revidert i 1993 og 2003. Første modell som kun fokuserte på programvareutvikling. Modellen støtter kun opp om selve utviklingsprosessen og ser bort fra kunde, marked og inntjening.

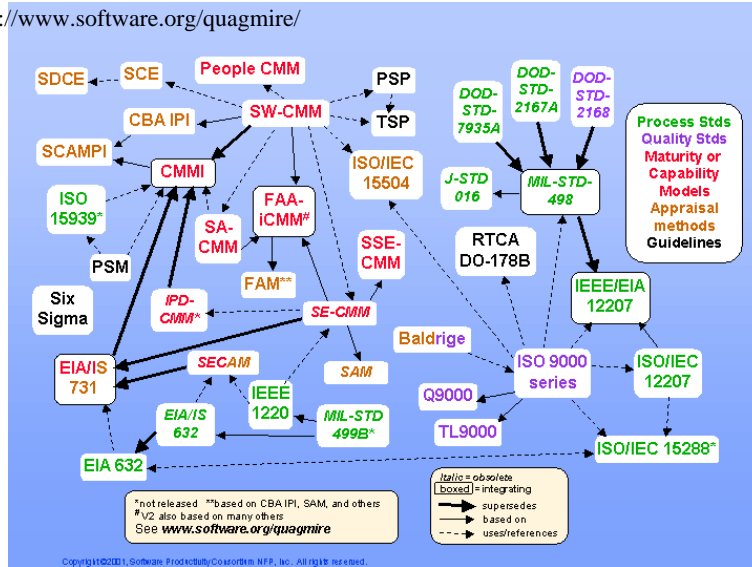
INF5180

34



"Frameworks quagmire"

<http://www.software.org/quagmire/>



35



Hva skal vi velge??

- Det blusser til tider opp "halvreligiøse kriger" om disse standardene. Her er noen utsagn:
 - "Alle er forskjellige og behovene er så spesielle at det ikke går å generalisere løsningen"
 - "Det fører bare til byråkrati"
 - "Det går på bekostning av kreativiteten"
 - "Det koster urimelig mye å sertifiseres/evalueres"
 - "Du må måle din egen ytelse opp mot dine egne målsetninger for å kunne bli bedre"
- I tillegg er det uenighet om den praksis som foreskrives er god/riktig praksis
- Noen mener det er unødvendig å bedrive systematisk prosessforbedring på den ene eller andre måten.
 - Øvelse: Hvilke argumenter finnes for dette synet?

INF5180

36